Abstract

Data-driven decision making is increasingly viewed as essential in a globally competitive society. Initiatives to augment standardized testing with performance-based assessment have increased as educators progressively respond to mandates for authentic measurement of student attainment. To meet this challenge, multidisciplinary rubrics were developed as a method of scoring student work samples. The current study utilized confirmatory factor analysis to examine ratings of student work (N = 245) using the Quantitative Literacy VALUE Rubric from the Association of American Colleges and Universities. The study examined a conceptual model of the six skill measures from the rubric to validate whether, taken together, they are reliable measures of a single general construct—Empirical and Quantitative Skill (EQS), a Texas Core Curriculum objective. The model confirmed that the six measures in the rubric (*Interpretation, Representation, Calculation, Application/Analysis, Assumptions, and Communication*) appeared to describe a single construct. Results support using the Quantitative Literacy VALUE Rubric for assessing EQS.

AUTHORS

Julie S. Gray, Ph.D.
The University of Texas at Arlington

Melissa A. Brown, M.Ed.
The University of Texas at Arlington

John P. Connolly, Ph.D.
The University of Texas at Arlington

Examining Construct Validity of the Quantitative Literacy VALUE Rubric in College-level STEM Assignments

An individual's quantitative literacy and competence with data evaluation is helpful in all areas of life, including academia. Because data-driven decision making is increasingly viewed as essential in a globally competitive society, educational objectives often emphasize learning outcome elements such as data analysis and how to use the data to draw conclusions. Data analysis without an understanding of the implications limits appropriate actions that can be taken by individuals and businesses (Tufte, 1997). Further, hiring managers seek individuals with empirical and quantitative skills because they have the ability to see connections and systemic problems (National Association of Colleges and Employers, 2016). Indeed, findings from the Spellings Commission panel stated that, "better data about real performance and lifelong working and learning ability is absolutely essential if we are to meet national needs and improve institutional performance" (U.S. Department of Education, 2006, p. 30).

Over the past two decades educational policies in the United States were changed by congressional legislation (e.g., No Child Left Behind Act [NCLB], 2001; Every Student Succeeds Act [ESSA], 2015). McGuinn (2006) maintains that the NCLB was implemented in response to public sentiment to hold educators accountable for the instruction students receive. More recently, initiatives to augment standardized testing with performance-based assessment (PBA) have increased as educators progressively respond to mandates for authentic measurement of student attainment. This progression is particularly reflected in the recently legislated ESSA (Gewertz, 2015), which is anticipated to go into full effect during the 2017–2018 academic year. The next section briefly reviews some of the policy implications for assessment professionals.

CORRESPONDENCE

Email
julie.gray@uta.edu

Impact of Policy Changes on Assessment Professionals

NCLB in particular affected the responsibilities of educational assessment professionals in requiring that each state must measure student progress for an academic

year using single summative tests (Gewertz, 2015). As a result, a reliance upon standardized tests quickly developed to assess student attainment and inform process improvements in educational service delivery (Supovitz, 2009). Such testing often took the form of high-stakes, multiple-choice examinations. However, in the last decade, initiatives to extend assessment methods to include performance-based student work have gained momentum at many institutions (State Higher Education Executive Officers Association [SHEEO], 2016). As ESSA implementation moves toward completion, assessment professionals and state officials anticipate that it will provide them with options that include multiple measures during an academic year, including merging results from both standardized tests and performance-based tests (Gewertz, 2015). While many call the assessment of performance-based work a more authentic method of rating student attainment (Montgomery, 2002; Peden, Reed, & Wolfe, 2017; Rhodes, 2010; Rhodes & Finley, 2014), efforts to validate the way it is rated or scored present challenges for educators (Montgomery, 2002).

Unlike standardized tests, PBAs typically consist of written student work samples (e.g., essays, experimental or research lab summaries, and presentations).

PBA Challenges

PBA implies that in response to the assignment prompt, a student reveals the skills they have attained to date. That is, the student response contains authentic agreement between what the student knows and their ability to demonstrate that knowledge (Cobb, 2014). Unlike standardized tests, PBAs typically consist of written student work samples (e.g., essays, experimental or research lab summaries, and presentations). However, while PBA holds an advantage of authenticity it also presents a disadvantage. Montgomery (2002) lists concerns reported in the literature, including the difficulty of avoiding rater subjectivity when scoring authentic student work samples.

In contrast, normed scores for standardized tests for specific student populations typically guide comparisons based on equity and excellence. PBA often requires the introduction of a rubric to increase rater objectivity. Use of rubrics adds structure and consistency to the performance level assessment and comparisons (Montgomery, 2002).

VALUE Rubric Development as a Solution-Oriented Assessment Approach

Indeed, Montgomery (2002) recommended the use of rubrics for assessing authentic student work because they are tools that communicate to students the expected elements to include in the completed assignment. Rubrics for setting criteria and determining student attainment of the target objectives have been suggested to uphold equity and excellence for all students (Montgomery, 2002; Peden et al., 2017). That said, an evidence-based approach for evaluating PBAs using validated rubrics was needed.

A campus-based assessment initiative, led by the Association of American Colleges and Universities (AAC&U), published 16 Valid Assessment of Learning in Undergraduate Education (VALUE) Rubrics (AAC&U, 2017b). Faculty and other educational professionals gathered from over 100 different institutions of higher education, under the direction of the AAC&U, to develop the rubrics. The VALUE rubrics were designed to be scoring guides that can be used by universities to evaluate authentic student work samples. Further, the AAC&U outlined four families of Essential Learning Outcomes in order to advance VALUE rubrics as relevant assessment tools across a wide range of disciplines, courses, and objectives (National Leadership Council for Liberal Education & America's Promise, 2008).

The VALUE rubrics were designed to be scoring guides that can be used by universities to evaluate authentic student work samples.

These VALUE rubrics serve as a scaffold to government policies that endeavor to guarantee the quality of education across the United States for all students (AAC&U, 2017b). Though policies vary by state, they broadly included six educational objectives: critical thinking, communication, empirical and quantitative skill (EQS), teamwork, social responsibility, and personal responsibility. In the southwestern United States, the Texas Higher Education Coordinating Board (THECB) adopted the six aforementioned objectives for implementation in the most recent revision of the Texas Core Curriculum (TCC; THECB, 2011). The THECB required that all two-year and four-year educational institutions submit regular reports detailing the assessment practices and results for student TCC objective attainment within general education courses that have been approved and designated for inclusion in the TCC (THECB, 2011). Decision making regarding methodologies for rating these performance-based student work samples was left to the discretion of each institution by the THECB.

Preliminary studies supported by SHEEO and AAC&U consisted of a collaborative effort by 60 institutions in nine states who agreed to test the utility of the VALUE Rubric to rate authentic student work (SHEEO, 2016). In 2014–2015, they examined faculty ratings of authentic student work to determine levels that indicate healthy thresholds for student mastery (Lederman, 2015). While the multi-state collaborative vetted two rubrics in the practice of evaluating student work during its initial phase and current studies expanded to include more institutions, to date, they have not explored rubric construct validity. Studies are needed to investigate the extent to which the measures within each VALUE Rubric accurately represent a single construct.

Importance of Studying the Quantitative Literacy VALUE Rubric

Studies are needed to investigate the extent to which the measures within each VALUE Rubric accurately represent a single construct.

Case studies document the use of the VALUE Rubrics nationwide (AAC&U, 2017a; Peden et al., 2017). This study examines the construct validity of the AAC&U Quantitative Literacy VALUE Rubric for evaluating EQS, a TCC objective. EQS allows an individual to understand information or raw data that is presented in tables, charts, graphs, or figures and evaluate it to draw accurate conclusions. Identifying applications of EQS across academic disciplines is straightforward. The ability to take information, analyze it, and predict outcomes is a common theme in the hard sciences such as engineering, physics, chemistry, and biology. In addition, EQS is utilized across disciplines, for instance, in nursing, business, and psychology.

Individuals possessing skills such as EQS are in high demand because they can use this expertise to find evidence-based solutions. EQS is typically described using action verbs including identify, extract, validate, and report (Georgesen, 2015). Further, the process often follows an ordered set of action steps. For example, Georgesen (2015) extended the list as a set of four steps: 1) define, scope, identify, document; 2) extract, aggregate, transform, create; 3) develop, analyze, simulate, validate; and 4) report, recommend, implement, monitor. The extent to which these verbs can be translated into observable measures is essential to evaluating student attainment of the TCC objective EQS.

The current study focused on the measures within the Quantitative Literacy VALUE Rubric and its utility for measuring EQS. The six skill indicators measured by this rubric are *Interpretation, Representation, Calculation, Application/Analysis, Assumptions, and Communication*. Explanations for each are contained in the rubric (see Appendix). Our hypothesis is that there is a single underlying trait or "latent variable" of which the six different skills are indicators. In short, we wish to validate that the six different skills being assessed, taken together, are reliable measures of something more general.

Method

The current study focused on the measures within the Quantitative Literacy VALUE Rubric and its utility for measuring EQS.

The skills within the Quantitative Literacy VALUE Rubric were assessed using written samples of undergraduate student work from approved Signature Assignments embedded in the existing undergraduate TCC courses at a four-year public institution in an urban setting. The institution met requirements to serve as a Hispanic Serving Institution by the U.S. Department of Education (2016) and, importantly, received the R-1 designation by the Carnegie Classification of Institutions of Higher Education (2015), the definitive list for top doctoral research. The measurement of student attainment of EQS is of extreme interest because of the institutional focus on research.

Signature Assignments were designed to be authentic performance-based work in which students responded to pedagogically relevant prompts. For example, some Signature Assignments consisted of written summaries of actual lab experiments conducted by students in life and physical sciences courses. These papers, illustrated by tables and figures, essentially included measurable elements of *Interpretation, Representation, Calculation, Application/Analysis, Assumptions, and Communication*. All the Signature Assignments in this sample were collected from courses related to science, technology, engineering and math (STEM). Trained faculty and staff who participated in calibration and training exercises (described in more detail to follow) performed the ratings.

Participants

Signature Assignments were obtained from 296 undergraduates enrolled in core curriculum courses in STEM areas at the university. The readability of a portion of the assignments (n = 51) was poor because they were scanned copies of handwritten summaries from lab books or "blue books." As such, these 51 Signature Assignments were dropped from the sample and not rated. Ratings were available for 245 of the student Signature Assignments. Over half of the participants were female (61%; n = 149), which closely matched the gender ratio at the university. The sample also reflected a rich diversity of students. About a third of the student participants identified as White (33%; n = 80), almost a third identified as Hispanic (27%; n = 67), and the balance was split between African American; Asian; foreign, nonresident alien; multiple ethnicity; and unknown, not specified. Students represented nine of ten colleges and schools at the university (see Table 1).

Table 1. *Student Characteristics for the Rated Sample of Signature Assignments*

Categorical Variables	*N*	%
Gender		
Female	149	60.8
Male	96	39.2
Ethnicity		
African American	33	13.5
Asian	49	20.0
Caucasian	80	32.7
Foreign, nonresident alien	6	2.4
Hispanic	67	27.3
Multiple	5	2.0
Unknown, not specified	5	2.0
College/School		
College of Architecture	2	0.8
College of Business	24	9.8
College of Education	13	5.3
College of Engineering	15	6.1
College of Liberal Arts	26	10.6
College of Nursing	61	24.9
College of Science	61	24.9
School of Social Work	15	6.1
Undeclared	26	10.6
[a]Missing college or school information	2	0.8
Level		
Freshman	67	27.3
Sophomore	85	34.7
Junior	49	20.0
Senior and above	42	17.1
[a]Missing level information	2	0.8

Note: N = 245 for each of the categorical variable. [a] Information was missing

Procedure

Faculty currently teaching undergraduate courses in STEM areas agreed to submit the course set of authentic student work deemed as the Signature Assignment for this study. The syllabus for each core curriculum class at the university describes the Signature Assignment and the students enrolled in these courses complete it as they would any other assignment or required course work. The samples submitted for this assessment process were ungraded, de-identified copies. Steps to redact personal and academic information were followed for two reasons. The first was to prevent any bias among rater scores in response to the grade the paper received from the professor. The second was to protect the confidentiality of student, faculty, and course information.

Assessment Instrument

The Signature Assignments were assessed using the VALUE Rubric for Quantitative Literacy (AAC&U, 2009), which categorizes EQS into six measures: *Interpretation, Representation, Calculation, Application/Analysis, Assumptions, and Communication*. The rubric describes each measure and uses a four-point Likert scale for determining scores (see Appendix). Higher values indicate more evidence of EQS. Using the rubric, raters assigned a score to each of the six skill measures.

Typically, in student samples, the six measures are adequately represented in the narrative of the Signature Assignment. It is important to note that visual communication in the form of charts, graphs, and figures enhanced the identification of the *Representation and Communication* measures. This is not unexpected because communication (written and visual) is required for fleshing out and articulating ideas in STEM areas. Visual communication is particularly important, and in many cases essential, for depicting information in STEM areas.

Raters, Rater Calibration, and Scoring

EQS allows an individual to understand information or raw data that is presented in tables, charts, graphs, or figures and evaluate it to draw accurate conclusions.

For the purposes of this study, the unit of analysis was an individual rater's score for a particular Signature Assignment. Raters scored the student writing samples during a scheduled scoring day so each paper was read and then rated by at least two separate raters working independently in a group setting. The rater group included ten faculty members and professional staff with advanced degrees. Scoring day began with an orientation and description of the rating process. Then, the entire group read one anchor paper chosen by the facilitator. Next, the facilitator led a discussion focused on reaching a common understanding of the EQS measures and finding exemplar indicators within the anchor paper for the rubric's levels of mastery. Then the rating process began and raters individually read their assigned papers to score each measure with the rubric (four-point Likert scale). Two raters independently rated each paper. Measure scores were calculated as the average of both scores. The facilitator checked each paper, after the completion of the two ratings, to review whether disagreement between measure ratings exceeded acceptable metrics. If so, the facilitator assigned a third rater as a separate, impartial mediator. In those cases (n = 4) the outlier of the three ratings was replaced.

Inter-rater Agreement

To examine the agreement between raters, an estimate of inter-rater reliability was calculated to see how frequently the rater pairs agreed on the score when rating the same paper. Conclusions about the consistent measurement of the six measures depend on this estimate. A calculation of the intraclass correlation coefficient (ICC) was used to determine the level of inter-rater agreement. High ICC values indicate more agreement between raters. A one-way random model was used to measure consistency within the mean measure values. ICC values for *Interpretation, Representation, Calculation, Application/Analysis, Assumptions, and Communication* indicated good inter-rater agreement (see Table 2) even though rater pairs varied across ratings, which typically results in lower ICC values (Landers, 2015).

Table 2. *ICC Values by Measure*

Measure	ICC Value
Interpretation	.52
Representation	.51
Calculation	.47
Application/Analysis	.56
Assumptions	.51
Communication	.60

Note: N = 245 for each measure.

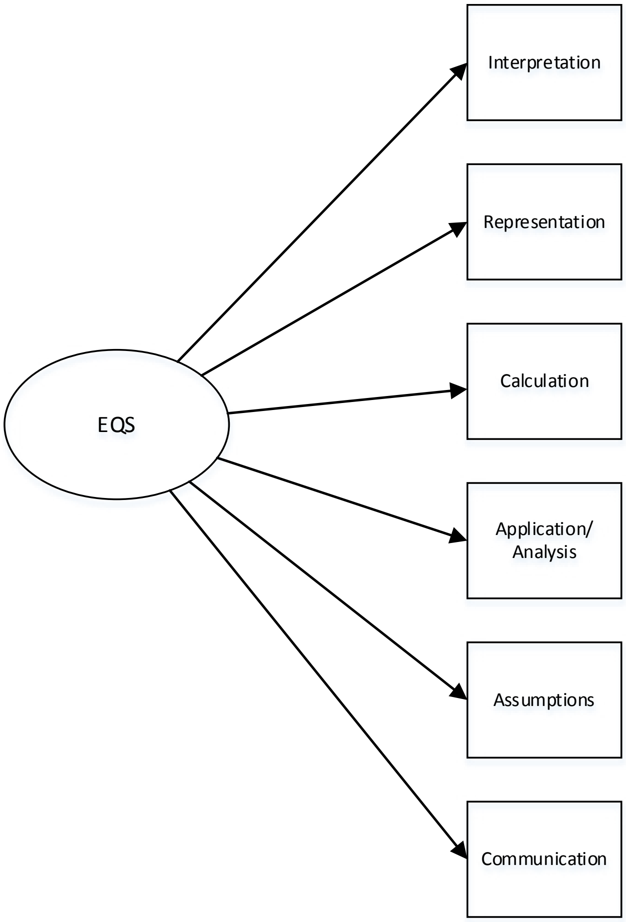

Figure 1. Conceptual Model of Underlying EQS Traits

Analysis Plan

We used confirmatory factor analysis to assess whether the six measured skills are reliable indicators of an underlying more general construct (Brown 2006). One key advantage of this approach is the ability to isolate the underlying construct from random error variance in the indicator measures. Further, correlations across the error components of each survey item can also be modeled to account for method effects that detract from the underlying construct, such as any tendency to rate two of the skills more similarly than the others. Figure 1 depicts the conceptual model (H_0).

Signature Assignments were designed to be authentic performance-based work in which students responded to pedagogically relevant prompts.

Because the measure ratings are in the form of a Likert scale, and therefore categorical, we used a mean- and variance-adjusted weighted least squares (WLSMV) estimator to estimate the loadings of each measure on the underlying EQS trait (Muthén & Muthén, 1998–2012).

The same estimator also yields fit statistics that provide information on the overall reliability of the model in terms of its ability to reproduce the variances and covariances of the indicator measures. Ideally, the model reports a nonsignificant chi square value indicating that imposing the hypothesized structure on the data does not amount to a substantial loss of information. However, since chi-square statistics are proportional to sample size other statistics are commonly used to assess model fit. In particular, a Root Mean Square Error of Approximation (RMSEA) statistic that is below 0.05 and a Comparative Fit Index (CFI) greater than 0.95 indicates a model that is a good fit to the data (Byrne, 2012).

Results

All the analyses were conducted in Mplus v.7.31 (Muthén & Muthén, 2012), which also reports ways of improving the model via modification indices. Analysis of the set of ratings from rater 1 and then the set from rater 2 (from the rater parings) indicated that

Current efforts toward the use of PBA to augment standardized testing with students present a challenge for educators because of the possible rater bias and other differences in scoring authentic student work; thus, there is a need to validate the rubrics that raters use.

significant model improvement would be obtained by allowing the random error variances in the *Representation and Calculation* measures to correlate. The fit statistics of the two models, i.e., the model with the specified error correlation (the H_1 model) and the model with no error correlations (the H_0 model), are summarized in Table 3. The H_1 model met all the criteria of a well-fitting model in terms of the key fit statistics: chi square, RMSEA, and CFI. The table also showed a significant loss of fit for the H_0 model in terms of a chi-square difference test.

The unstandardized loadings of each of the six skill measures on the underlying EQS latent variable are summarized in Table 4. The standardized estimates, along with associated standard errors, are shown in Figure 2. Also included in Figure 2 is the estimate for the error correlation between *Representation and Calculation*.

Table 3. *Model fit statistics for the H_1 and H_0 models with X^2 difference test*

	N	χ^2	*df*	P-Value	RMSEA	CFI
H_1 Model	245	9.31	8	0.317	0.03	0.99
H_0 Model	245	47.31	9	0.000	0.13	0.98
Difference Test		19.40	1	0.000		

The estimates in Table 4 are akin to regression estimates of the effect of the underlying EQS trait on the skill in question—all of which were statistically significant at the 0.01 alpha level. The three strongest indicators were *Communication, Application/Analysis, and Interpretation*, and the amount of variance in these indicators explained by EQS is 77%, 73%, and 71%, respectively. Weaker effects were found in the case of *Calculation* (53%), *Assumptions* (47%), and *Representation* (42%).

Table 4. *Weighted Least Squares estimates for the six skill measures*

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Interpretation[a]	1.00	0.00		
Representation	0.77	0.07	11.20	0.000
Calculation	0.87	0.07	12.72	0.000
Application/Analysis	1.02	0.06	17.62	0.000
Assumptions	0.81	0.06	12.91	0.000
Communication	1.04	0.06	18.98	0.000

Note: [a]For the purpose of scaling the latent ariable, *Interpretation* is treated as the marker indicator. As such, the associated loading of this indicator on EQS is set at a value of one (Brown, 2006, p.71).

Discussion

Current efforts toward the use of PBA to augment standardized testing with students present a challenge for educators because of the possible rater bias and other differences in scoring authentic student work; thus, there is a need to validate the rubrics that raters use. The goal of the current study was to examine the construct validity of the Quantitative Literacy VALUE Rubric, one of 16 rubrics developed by the AAC&U. Overall, the findings show that the six underlying skill measures tapped into a common underlying EQS trait. These results extend previous research that has primarily focused on the use of the rubrics to study trends in student attainment (SHEEO, 2016).

Summary of Findings

Our hypothesis-testing results suggested that the six measures each reflect EQS as an underlying trait and that raters using the rubric produced valid EQS scores. Significant consistency was confirmed by analyzing rubric ratings of authentic student work from

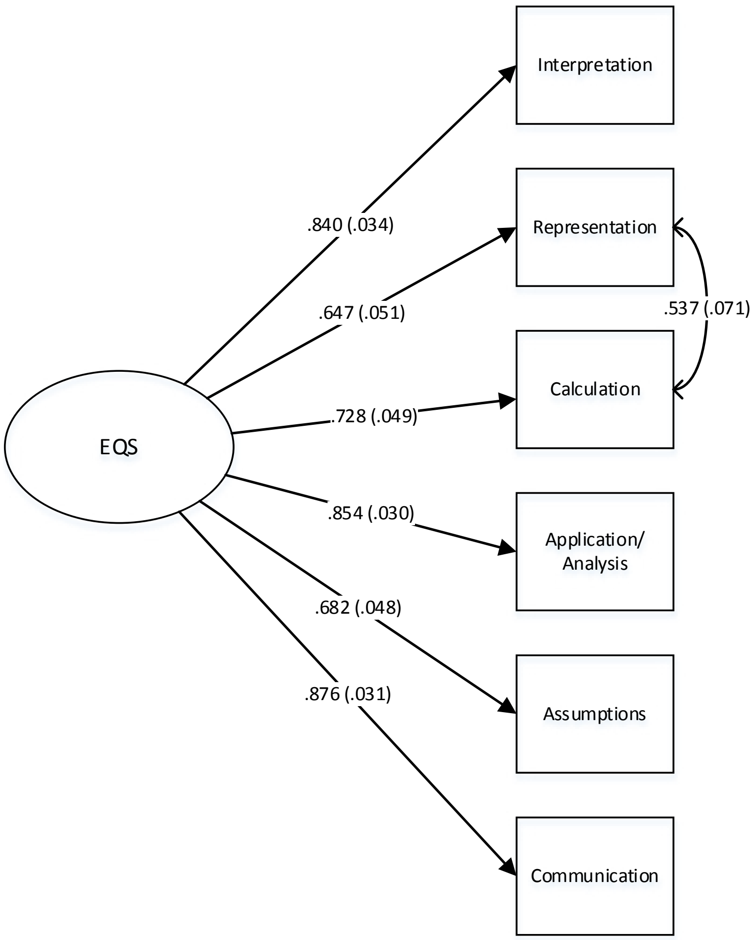

Figure 2. Standardized estimates for final solution (H_1)

STEM courses at a four-year public university. Each of the six measured skills loaded on the same construct and the model accounted for a large proportion of variance in each of the indicators. This validates that the six different skills assessed by the Quantitative Literacy VALUE Rubric are reliable measures of the general trait, EQS. Though the importance of nonsubjective measures of PBA has been well established, to our knowledge this is the first study that confirmed how well the measured skills in the Quantitative Literacy VALUE Rubric fit together as a model of EQS.

The goal of the current study was to examine the construct validity of the Quantitative Literacy VALUE Rubric, one of 16 rubrics developed by the AAC&U.

In addition, the pattern of results indicated three measures with very strong contributions to the model, *Interpretation, Application/Analysis, and Communication*. These three skill measures are widely used in statistical texts to describe the analytical process researchers use after research questions are posed, studies are designed, and data are collected. Without them, the research process is just a collection of numbers, and does not contribute answers to research questions that often have real consequence in many fields. Indeed, national surveys of employers repeatedly list skills involving *Interpretation, Application, and Communication* as essential qualities in job applicants (National Association of Colleges and Employers, 2016). The model confirmed the strength of the rubric in representing these highly marketable skills—those that are involved in quantitative literacy.

In further support for the model, analyses revealed inter-rater reliability estimates in the moderate to good range for the six measures. This suggests that rater calibration activities conducted on scoring day may have held a degree of utility in terms of promoting agreement among raters. The literature about VALUE rubrics contains many case studies of the use of calibration as a best practice (AAC&U, 2017a; Finley, 2011; Peden et al., 2017) yet, to our knowledge, it does not contain findings related to calibration activity effectiveness that directly

compared a trained group of raters with a group that did not undergo any sort of training.

This validates that the six different skills assessed by the Quantitative Literacy VALUE Rubric are reliable measures of the general trait, EQS.

In addition, while inter-rater agreement may have differed with the introduction of more than two raters for all Signature Assignments, the study design accounted for the importance of good inter-rater agreement by planning the facilitator-led calibration activities and using a third rater to mediate unacceptable differences. Indeed, Stanny, Gonzalez, and McGowan (2015) mention improvement in rater agreement through the use of similar activities that operationalize rubric guidelines with "notes [added to the rubric] about difficult decisions, to build and maintain consensus for future decisions" (p. 905). Further, Finley (2011) recommends that rating sessions include the type of facilitator-led discussions that were used in this study before the application of the rubric to ensure adequate agreement. Though not a primary focus of the current study, findings suggested that the level of agreement for the ratings in the sample provided adequate justification for proceeding with the analysis of the rubric's construct validity.

In addition to strengths already mentioned, the model improved when the association between *Calculation* and *Representation* was allowed to covary. This makes sense because a single-minded focus on *Calculation* makes drawing conclusions hard to visualize and a skill such as *Representation* strengthens its meaning. In that way, *Calculation* and *Representation* dovetail together. In practice, calculation turns to representation to derive meaning and understanding as two parts of the same whole. In the process of problem solving, making a visual representation is a natural process for deriving meaning from computational problems (Van Garderen & Montague, 2003) and for enhancing the decision-making value of quantitative information (Tufte, 1997).

Continued efforts are needed to promote the use of authentic student work in educational assessment.

Limitations

The findings of the current study are promising but a few limitations should be noted. For instance, student samples only represented STEM courses in the life and physical sciences. This limited the ability to examine the independent effects of other types of courses and potential confounds. In future studies, course types should be extended to include all three of the foundational component areas required by the THECB (life and physical science, mathematics, and social and behavioral science). Though all students at the university were also required to take courses across eight foundational component areas as part of the TCC, conclusions would be strengthened through the incorporation of a wider range of courses. Additionally, performance-based work was gathered only from TCC-approved courses and the naturalistic design of the study did not allow for randomized assignment of papers from across all the STEM courses on campus regardless of level. Nonetheless, the student demographics suggest that the sample was consistent with the campus population as a whole.

Conclusion

Continued efforts are needed to promote the use of authentic student work in educational assessment. This study examined a widely utilized rubric using a relatively large sample of STEM assignments to capitalize on the strength of the AAC&U initiatives that measure student attainment of broadly accepted educational learning objectives. Results suggest that the six skill measures contained in the Quantitative Literacy VALUE Rubric fit together well to explain EQS. Consequently, efforts to promote VALUE rubrics have the potential to accurately measure student attainment of EQS. Further research is needed to confirm the construct validity of the full array of AAC&U VALUE Rubrics. Continuation of this line of inquiry is essential for maximizing the effectiveness of PBA.

Keywords: quantitative literacy, empirical and quantitative skill, VALUE rubric, STEM, EQS, performance-based assessment, Texas Core Curriculum, AAC&U

Appendix

QUANTITATIVE LITERACY VALUE RUBRIC

for more information, please contact value@aacu.org

Association of American Colleges and Universities

Definition

Quantitative Literacy (QL) – also known as Numeracy or Quantitative Reasoning (QR) – is a "habit of mind," competency, and comfort in working with numerical data. Individuals with strong QL skills possess the ability to reason and solve quantitative problems from a wide array of authentic contexts and everyday life situations. They understand and can create sophisticated arguments supported by quantitative evidence and they can clearly communicate those arguments in a variety of formats (using words, tables, graphs, mathematical equations, etc., as appropriate).

Evaluators are encouraged to assign a zero to any work sample or collection of work that does not meet benchmark (cell one) level performance.

	Capstone 4	Milestones 3	2	1
Interpretation *Ability to explain information presented in mathematical forms (e.g., equations, graphs, diagrams, tables, words)*	Provides accurate explanations of information presented in mathematical forms. Makes appropriate inferences based on that information. *For example, accurately explains the trend data shown in a graph and makes reasonable predictions regarding what the data suggest about future events.*	Provides accurate explanations of information presented in mathematical forms. *For instance, accurately explains the trend data shown in a graph.*	Provides somewhat accurate explanations of information presented in mathematical forms, but occasionally makes minor errors related to computations or units. *For instance, accurately explains trend data shown in a graph, but may miscalculate the slope of the trend line.*	Attempts to explain information presented in mathematical forms, but draws incorrect conclusions about what the information means. *For example, attempts to explain the trend data shown in a graph, but will frequently misinterpret the nature of that trend, perhaps by confusing positive and negative trends.*
Representation *Ability to convert relevant information into various mathematical forms (e.g., equations, graphs, diagrams, tables, words)*	Skillfully converts relevant information into an insightful mathematical portrayal in a way that contributes to a further or deeper understanding.	Competently converts relevant information into an appropriate and desired mathematical portrayal.	Completes conversion of information but resulting mathematical portrayal is only partially appropriate or accurate.	Completes conversion of information but resulting mathematical portrayal is inappropriate or inaccurate.
Calculation	Calculations attempted are essentially all successful and sufficiently comprehensive to solve the problem. Calculations are also presented elegantly (clearly, concisely, etc.)	Calculations attempted are essentially all successful and sufficiently comprehensive to solve the problem.	Calculations attempted are either unsuccessful or represent only a portion of the calculations required to comprehensively solve the problem.	Calculations are attempted but are both unsuccessful and are not comprehensive.
Application / Analysis *Ability to make judgments and draw appropriate conclusions based on the quantitative analysis of data, while recognizing the limits of this analysis*	Uses the quantitative analysis of data as the basis for deep and thoughtful judgments, drawing insightful, carefully qualified conclusions from this work.	Uses the quantitative analysis of data as the basis for competent judgments, drawing reasonable and appropriately qualified conclusions from this work.	Uses the quantitative analysis of data as the basis for workmanlike (without inspiration or nuance, ordinary) judgments, drawing plausible conclusions from this work.	Uses the quantitative analysis of data as the basis for tentative, basic judgments, although is hesitant or uncertain about drawing conclusions from this work.
Assumptions *Ability to make and evaluate important assumptions in estimation, modeling, and data analysis*	Explicitly describes assumptions and provides compelling rationale for why each assumption is appropriate. Shows awareness that confidence in final conclusions is limited by the accuracy of the assumptions.	Explicitly describes assumptions and provides compelling rationale for why assumptions are appropriate.	Explicitly describes assumptions.	Attempts to describe assumptions.
Communication *Expressing quantitative evidence in support of the argument or purpose of the work (in terms of what evidence is used and how it is formatted, presented, and contextualized)*	Uses quantitative information in connection with the argument or purpose of the work, presents it in an effective format, and explicates it with consistently high quality.	Uses quantitative information in connection with the argument or purpose of the work, though data may be presented in a less than completely effective format or some parts of the explication may be uneven.	Uses quantitative information, but does not effectively connect it to the argument or purpose of the work.	Presents an argument for which quantitative evidence is pertinent, but does not provide adequate explicit numerical support. (May use quasi-quantitative words such as "many," "few," "increasing," "small," and the like in place of actual quantities.)

References

Association of American Colleges and Universities. (2009). *Quantitative literacy VALUE rubric*. Retrieved from https://www.aacu.org/value/rubrics/quantitative literacy

Association of American Colleges and Universities. (2017a). *Campus models and case studies*. Retrieved from https://www.aacu.org/campus-model/3305

Association of American Colleges and Universities. (2017b). *VALUE rubrics*. Retrieved from https://www.aacu.org/value-rubrics

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York and London: The Guilford Press.

Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications and programming*. New York and London: Routledge.

Carnegie Classification of Institutions of Higher Education. (2015). *Classification update: List of R-1 doctoral universities*. Retrieved from http://carnegieclassifications.iu.edu

Cobb, R. (2014). *The paradox of authenticity in a globalized world*. New York: Palgrave Macmillan.

Finley, A. P. (2011). How reliable are the VALUE rubrics? *Peer Review, (14)*1, 31–33.

Georgesen, J. (2015). *Evolving from big data to smart data: New ways CX researchers predict customer behavior*. Retrieved from http://mrweek.com/content

Gewertz, C. (2015). ESSA's flexibility on assessment elicits qualms from testing experts. *Education Week*. Retrieved from http://www.edweek.org/ew/articles/2015/12/21/essas-flexibility-on-assessment-elicits-qualms-from.html

Landers, R. N. (2015). Computing intraclass correlations (ICC) as estimates of interrater reliability in SPSS. *The Winnower*. doi:10.15200/winn.143518.81744

Lederman, D. (2015). New effort aims to standardize faculty-driven review of student work. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2015/09/25/new-effort-aims-standardize-faculty-drivenreview-student-work

McGuinn, P. J. (2006). *No Child Left Behind and the transformation of federal education policy, 1965–2005*. Lawrence, KS: University Press of Kansas.

Montgomery, K. (2002). Authentic tasks and rubrics: going beyond traditional assessments in college teaching. *College Teaching, (50)*1, 34–40. doi:10.1080/87567550209595870

Muthén, L.K. & Muthén, B.O. (1998-2012). *Mplus user's guide*. (7th ed.). Los Angeles, CA: Author.

National Association of Colleges and Employers. (2016). *Job Outlook 2016*. Bethlehem, PA.

National Leadership Council for Liberal Education & America's Promise. (2008). *College learning for the new global century*. Washington, DC: Association of American Colleges and Universities.

Peden, W., Reed, S., & Wolfe, K. (2017). *Rising to the LEAP challenge: Case studies of integrative pathways to student work*. Washington, DC: Association of American Colleges and Universities.

Rhodes, T. (Ed.). (2010). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Washington, DC: Association of American Colleges and Universities.

Rhodes, T. & Finley, A. (2014). *The VALUE rubrics: Frequently asked questions about development, interpretation, and use of rubrics on campuses*. Retrieved from http://www.sheeo.org/sites/default/files/project-files/VALUERubrics_Webinar %28R%29.pptx

Stanny, C., Gonzalez, M., & McGowan, B. (2015). Assessing the culture of teaching and learning through a syllabus review. *Assessment & Evaluation in Higher Education, 40*(7), 898–913. doi:10.1080/02602938.2014.956684

State Higher Education Executive Officers Association. (2016). *MSC: A multi-state collaborative to advance learning outcomes assessment*. Retrieved from http://www.sheeo.org/projects/msc-multi-state-collaborative-advance-learning-outcomes-assessment#

Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change, 10*(2–3), 211–227. Retrieved from http://link.springer.com/journal/10833

Texas Higher Education Coordinating Board. (2011). *Texas Core Curriculum*. Retrieved from http://www.thecb.state.tx.us/index.cfm?objectid=417252EA-B240-62F79F6A1A125C83BE08

Tufte, E. R. (1997). *Visual explanations: images and quantities, evidence and narrative*. Cheshire, CT: Graphics Press.

U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education*. Washington, D. C.

U.S. Department of Education. (2016). *FY 2016 eligible Hispanic-serving institutions*. Retrieved from https://www2.ed.gov/about/offices/list/ope/idues/hsi-eligibles-2016.pdf

Van Garderen, D., & Montague, M. (2003). Visual-spatial representation, mathematical problem solving, and students of varying abilities. *Learning Disabilities Research & Practice, 18*(4), 246. doi:10.1111/1540-5826.00079